

Plug-and-Play Diffusion Features for Text-Driven Image-to-Image Translation

Year: 2022 | Citations: 862 | Authors: Narek Tumanyan, Michal Geyer, Shai Bagon

Abstract

Large-scale text-to-image generative models have been a revolutionary breakthrough in the evolution of generative AI, synthesizing diverse images with highly complex visual concepts. However, a pivotal challenge in leveraging such models for real-world content creation is providing users with control over the generated content. In this paper, we present a new framework that takes text-to-image synthesis to the realm of image-to-image translation - given a guidance image and a target text prompt as input, our method harnesses the power of a pre-trained text-to-image diffusion model to generate a new image that complies with the target text, while preserving the semantic layout of the guidance image. Specifically, we observe and empirically demonstrate that fine-grained control over the generated structure can be achieved by manipulating spatial features and their self-attention inside the model. This results in a simple and effective approach, where features extracted from the guidance image are directly injected into the generation process of the translated image, requiring no training or fine-tuning. We demonstrate high-quality results on versatile text-guided image translation tasks, including translating sketches, rough drawings and animations into realistic images, changing the class and appearance of objects in a given image, and modifying global qualities such as lighting and color.